

The relative performance characteristics of the logistic European System for Cardiac Operative Risk Evaluation score and the Society of Thoracic Surgeons score in the Placement of Aortic Transcatheter Valves trial

Nirat Beohar, MD, FACC, FSCAI,^a Brian Whisenant, MD,^b Ajay J. Kirtane, MD, SM,^c Martin B. Leon, MD,^c E. Murat Tuzcu, MD,^d Raj Makkar, MD,^e Lars G. Svensson, MD, PhD,^d D. Craig Miller, MD,^f Craig R. Smith, MD,^c Augusto D. Pichard, MD,^g Howard C. Herrmann, MD,^h Vinod H. Thourani, MD,ⁱ Wilson Y. Szeto, MD,^h Scott Lim, MD,^j Michael Fischbein, MD,^f William F. Fearon, MD,^f William O'Neill, MD,^k Ke Xu, PhD,^c Todd Dewey, MD,^l and Michael Mack, MD^l

Objectives: The logistic European System for Cardiac Operative Risk Evaluation (LES) score and the Society of Thoracic Surgeons (STS) score are validated to predict 30-day outcomes following surgical aortic valve replacement (SAVR) with or without coronary artery bypass grafting. Their performance when applied to patients undergoing transcatheter aortic valve replacement (TAVR) is controversial.

Methods: We compared predicted and observed 30-day/in-hospital and 1-year mortality of patients undergoing TAVR in the first Placement of Aortic Transcatheter Valves trial and continued access registry (N = 2466). The performance of the LES and STS scores (prospectively calculated) was evaluated using standard assessments of discrimination and calibration. Performance of STS and LES scores among 307 patients undergoing SAVR from the high-risk cohort of the randomized trial were also examined.

Results: In patients undergoing TAVR, the observed 30-day/in-hospital mortality was 6.5%, whereas the predicted 30-day mortality was higher by both STS score (11.4% ± 3.9%) and LES score (26.6% ± 16.2%). The discrimination for both scores was poor for 30-day/in-hospital and 1-year mortality. Calibration was better for STS score than for LES at 1 year but poor for both at 30 days among TAVR cohort. These results were consistent among the subgroups of patients undergoing transfemoral and transapical access; however, the STS score had better performance among the high-risk patients who underwent SAVR at 30 days but not 1 year.

Conclusions: The STS and LES surgical risk scores overestimated 30-day/in-hospital mortality and were poor discriminators of post-TAVR mortality, but the calibration of the STS score was better in these high-risk patients. These data highlight the need for TAVR-specific risk models to optimize patient selection. (J Thorac Cardiovasc Surg 2014;148:2830-7)

From the Cardiac Catheterization Laboratory,^a Columbia University Division of Cardiology, Mount Sinai Medical Center, Miami Beach, Fla; Intermountain Medical Center,^b Salt Lake City, Utah; Columbia University Medical Center/New York Presbyterian Hospital,^c New York, NY; Cleveland Clinic Foundation,^d Cleveland, Ohio; Cedars-Sinai Medical Center,^e Los Angeles, CA; Stanford University,^f Stanford, Calif; Medstar Washington Hospital Center,^g Washington, DC; Hospital of the University of Pennsylvania,^h Philadelphia, Pa; Emory University School of Medicine,ⁱ Atlanta, Ga; University of Virginia,^j Charlottesville, Va; Center for Structural Heart Disease,^k Henry Ford Healthcare System, Detroit, Mich; and Baylor Healthcare System,^l Dallas, Tex.

Clinical trial registration: clinicaltrials.gov No. NCT00530894.

The PARTNER trial was funded by Edwards Lifesciences and designed collaboratively by the steering committee and the sponsor. Our analysis was carried out by academic investigators under the auspices of the PARTNER Publications Office with no additional involvement of the sponsor.

Disclosures: B.W. received consulting fees from Edwards Lifesciences; M.B.L., E.M.T., L. S., and M.M. received travel reimbursements from Edwards Lifesciences related to their work as unpaid members of the PARTNER trial executive committee; R.M. received research grant support from Edwards Lifesciences and St. Jude Medical, and consulting fees from Abbott Vascular, Cordis, and Medtronic; D.C.M. is supported by an R01 research grant from the NHLBI #HL67025,

received consulting fees from Abbott Vascular, St Jude Medical, and Medtronic, in addition to being an unpaid member of the PARTNER trial executive committee; A.D.P. received consulting fees from Edwards Lifesciences; H.H. received consulting fees from St Jude Medical and Paieon and has equity in MicroInterventional Devices; V.H.T. received consulting fees from Edwards Lifesciences, Sorin Medical, St Jude Medical, and DirectFlow; W.Y.S. received research grant support from Edwards Lifesciences and Medtronic and consulting fees from MicroInterventional Devices; S.L. received consulting fees from Boston Scientific Corporation, Guerbet, and St Jude Medical; W.F.F. collected travel reimbursements from Edwards Lifesciences related to his role on the steering committee of the PARTNER 2 trial; and W.O.N. received consulting fees from Medtronic. All other authors have nothing to disclose with regard to commercial support.

Received for publication Dec 23, 2013; revisions received March 17, 2014; accepted for publication April 4, 2014; available ahead of print May 10, 2014.

Address for reprints: Nirat Beohar, MD, FACC, FSCAI, Cardiac Catheterization Laboratory, Columbia University Division of Cardiology, Mount Sinai Medical Center, 4300 Alton Rd, Miami Beach, FL 33138 (E-mail: Nirat.Beohar@msmc.com).

0022-5223/\$36.00

Copyright © 2014 by The American Association for Thoracic Surgery

<http://dx.doi.org/10.1016/j.jtcvs.2014.04.006>

Abbreviations and Acronyms

AS	= aortic stenosis
LES	= logistic European System for Cardiac Operative Risk Evaluation
PARTNER	= Placement of Aortic Transcatheter Valve trial
STS	= Society of Thoracic Surgeons
SAVR	= surgical aortic valve replacement
TAVR	= transcatheter aortic valve replacement
TA	= transapical
TF	= transfemoral

Supplemental material is available online.

Patients considered for transcatheter aortic valve replacement (TAVR) for severe symptomatic aortic stenosis (AS) are frequently elderly, frail, and have multiple comorbidities.^{1,2} These comorbidities, individually or in combination, are associated with increased perioperative morbidity and mortality as well as diminished late survival. The Placement of Aortic Transcatheter Valve trials (PARTNER I and PARTNER II) demonstrated that TAVR decreased mortality on an absolute scale by 20% compared with standard therapy among patients with severe AS deemed unsuitable for surgical aortic valve replacement^{2,3} (SAVR) and was associated with similar rates of survival compared with SAVR among high-risk patients at 1, 2, and 3 years of follow-up.^{1,4} Despite these encouraging results, overall mortality following TAVR in high-risk patients is high (44.2% at 3 years).⁵ Thus, accurate prediction of procedural and subsequent outcomes is critical to inform decisions as patients are considered for TAVR.

The logistic European System for Cardiac Operative Risk Evaluation (LES) score and the Society of Thoracic Surgeons (STS) predicted risk of mortality score are validated risk scores predicting 30-day outcomes in patients undergoing SAVR with or without coronary artery bypass grafting.^{6,7} Both the STS score and the LES score accurately predict mortality in low-risk patients undergoing bypass surgery but have not been well validated in high-risk patients undergoing SAVR.⁸ Similarly, neither the STS score nor the LES algorithm was designed to predict outcomes among patients undergoing TAVR. Notwithstanding the absence of prospective validation of these scores, the calculation of surgical risk scores has become widely integrated into the evaluation of potential TAVR patients and is even a prerequisite for risk assessment mandated by both the US Food and Drug Administration and the Centers for Medicare and Medicaid Services.

In our analysis, we sought to study the performance characteristics of the STS and LES surgical risk scores at 30 days and 1 year among patients undergoing TAVR and SAVR in the PARTNER I randomized and continued access cohorts. While acknowledging that both the STS score and LES score were designed to predict 30-day outcomes we wanted to additionally test if baseline STS and LES surgical risk scores could also be predictive of 1-year all-cause observed mortality.

Specifically we hypothesized that both the STS and LES surgical risk scores would show poor calibration and discrimination regarding 30-day and 1-year all-cause mortality when applied to high-risk patients with severe AS undergoing TAVR or SAVR in the PARTNER I randomized and continued access cohorts.

METHODS**Study Cohort**

In this as-treated analysis of the PARTNER I trial, we combined the randomized and continued access cohorts from the PARTNER I high-risk (cohort A) and inoperable (cohort B) trials to provide a cumulative population of 2552 TAVR patients and 313 SAVR patients. The PARTNER trial design has been previously described and published.^{1,2} Briefly, the PARTNER I trial enrolled patients with severe symptomatic AS (aortic valve area ≤ 0.8 cm² plus a peak velocity ≥ 4 mps or a mean transaortic valve gradient ≥ 40 mm Hg). Patients were divided into 2 cohorts: those who were considered to be candidates for surgery despite the fact that they were at high surgical risk, as defined by an elevated STS risk score or by the presence of coexisting conditions that would be associated with a predicted risk of death by 30 days after surgery of at least 15% (cohort A); and those who were not considered to be suitable candidates for surgery because they had coexisting conditions that would be associated with a predicted probability of 50% or more of either death by 30 days after surgery or a serious irreversible condition (cohort B). Patients from cohort B with a suitable iliofemoral vessel were randomized to transfemoral (TF) TAVR with the Edwards-Sapien heart valve system (Edwards Lifesciences, Irvine, Calif) or to standard medical care. Patients enrolled in cohort A were randomized to TAVR (TF if iliofemoral vessels were suitable or transapical [TA] TAVR if not) or to conventional SAVR. The institutional review board at each participating site approved the study and all patients provided written informed consent.

Study Outcomes

The clinical outcomes analyzed included observed 30-day/in-hospital and 1-year all-cause mortality. The observed mortality was defined as adjudicated 30-day/in-hospital (all-cause mortality within 30 days of the index procedure or during the index hospitalization) and 1-year all-cause mortality following the index TAVR or SAVR procedure. An independent clinical events committee adjudicated all clinical events.

Statistical Analysis

Continuous variables are presented as mean (± 1 standard deviation) and compared by Student *t* test or presented as median (interquartile range) and compared by Mann-Whitney *U* test. Categorical variables are presented as frequencies and percentages and compared by χ^2 test.

The LES score and STS predicted mortality score were prospectively calculated with online calculators (<http://www.euroscore.org>; data set: 2.61, <http://www.riskcalc.sts.org>) by study investigators at participating sites and integrated into a central database. The relationship between

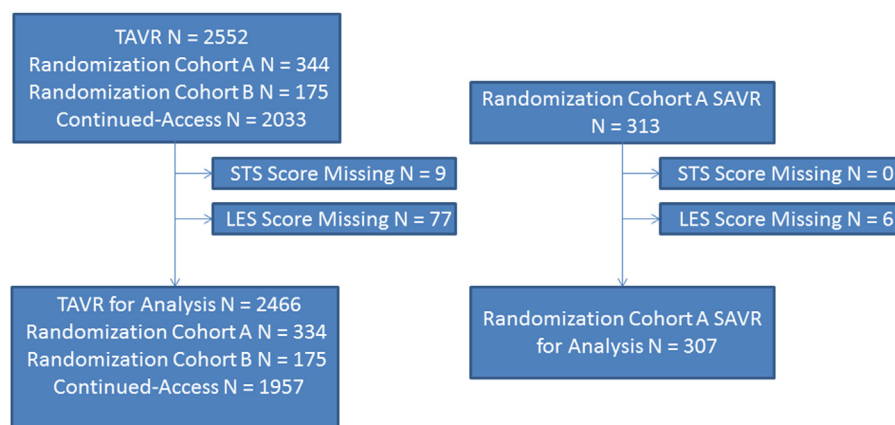


FIGURE 1. Flow of participants in transcatheter aortic valve replacement (TAVR) and surgical aortic valve replacement (SAVR) cohorts. STS, Society of Thoracic Surgeons; LES, logistic European System for Cardiac Operative Risk Evaluation.

LES and STS scores was quantified using Pearson correlation coefficient and linear regression analysis.

The performance of the LES and STS risk algorithms were evaluated in terms of their discrimination and calibration. Model discrimination was assessed using the area under the receiver operating characteristic (AUC) curve, analogous to the c statistic, with 95% confidence intervals. An AUC of 0.5 indicates no discriminatory ability, whereas an AUC of 1.0 represents perfect discrimination.^{9,10} To assess calibration, observed and predicted probabilities were graphically represented by a calibration plot. A smooth line was fitted and the calibration slope and intercept were reported. Perfect calibration was shown as a dashed line ($Y = X$ [45°] line, slope = 1, intercept = 0). Analyses of the performance characteristics of LES and STS scores when applied separately to patients undergoing TA and TF TAVR, cohort A (high-risk) and cohort B (inoperable) as well as those patients undergoing SAVR, were conducted. All statistical analyses were performed with SAS version 9.2 (SAS Institute Inc, Cary, NC).

RESULTS

A total of 2552 high-risk and inoperable patients underwent TAVR in cohort A ($n = 344$) and B ($n = 175$) of the randomized and continued access ($n = 2033$) PARTNER trials. Of these, STS score was missing for 7 patients and LES score was missing for 77 patients, resulting in a final cohort of 2466 patients included in this analysis; of 313 patients who were randomized to SAVR in cohort A of the randomized PARTNER trial 6 had missing LES data, which resulted in a final cohort of 307 patients who were also evaluated (Figure 1). The majority of patients were octogenarians, with an equal distribution of men and women undergoing TAVR, but there was higher prevalence of men in the SAVR group. The majority of patients undergoing TAVR or SAVR had an STS score $> 10\%$. Cardiovascular risk factors, cardiovascular conditions, and noncardiac conditions are detailed for both the TAVR and SAVR populations in Table 1.

Among patients undergoing TAVR, there were 165 of 2552 (6.5%) 30-day/in-hospital deaths. Of these, 111 of 2552 (4.3%) were adjudicated as due to cardiovascular

causes. At 1 year, all-cause mortality was 21.3% with 9.5% of deaths adjudicated as due to cardiovascular causes. Among 313 patients randomized to SAVR, there were 33 30-day/in-hospital deaths (10.5%), of which 11 (3.5%) were adjudicated as due to cardiovascular causes. At 1 year, all-cause mortality was 25.2% with 7.6% of deaths adjudicated as due to cardiovascular causes.

Risk Score Performance Characteristics of Patients Undergoing TAVR ($n = 2552$)

The mean LES score was more than 2 times higher than the mean STS score ($26.6\% \pm 6.2\%$ vs $11.4\% \pm 3.8\%$). The observed 30-day/in-hospital mortality was 6.5%; thus, the observed to predicted mortality ratio was 0.15 for the LES score and 0.57 for the STS score. There was a weak ($Y = 0.0828x + 9.1463$; $R = 0.3405$; $r^2 = 0.12$) relationship between the LES and STS scores.

Although the discrimination of STS score was better than LES score ($P = .024$), both were poor discriminators of 30-day/in-hospital outcomes (Table 2). When TA and TF access patients were analyzed separately, the 30-day discrimination for both STS score and LES score remained poor for both TA (LES AUC = 0.54 [95% CI, 0.48-0.60], STS score AUC = 0.59 [95% CI, 0.53-0.66]; $P = .23$) and TF (LES AUC = 0.51 [95% CI, 0.44-0.58], STS score AUC = 0.58 [95% CI, 0.49-0.66]; $P = .32$). No material difference in performance when stratified by enrollment cohort (high risk [cohort A] or inoperable [cohort B]) was noted. At 1 year, discrimination of both scores remained poor (Table 2).

Both risk scores demonstrated significant differences between predicted and observed mortalities throughout the range of mortality risk when analyzed at the 30-day/in-hospital time point (Table 3 and Figure 2, A and B). The predictive performance of both risk scores remained poor in the lower, mid, and high score predicted risk

TABLE 1. Baseline characteristics of the study population

	All TAVR	
	(cohort A+B+CA) (n = 2552)	SAVR (cohort A) (n = 313)
Age, y	85.64 (80.81-89.28)	84.94 (80.77-88.72)
Age ≥ 80 y	78.5% (1989/2534)	79.2% (248/313)
Male	52.4% (1336/2550)	57.2% (179/313)
Body mass index	25.74 (22.620-29.64)	25.69 (23.04-29.52)
Society of Thoracic Surgeons score	11.43 ± 3.83 (2543)	11.70 ± 3.37 (313)
Logistic European System for Cardiac Operative Risk Evaluation score	26.55 ± 16.25 (2468)	29.22 ± 15.13 (307)
Cardiovascular risk factors		
Any diabetes	37.2% (949/2550)	41.5% (130/313)
Hyperlipidemia	83.7% (2134/2550)	84.0% (263/313)
Smoking	48.4% (1233/2550)	49.8% (156/313)
Hypertension	91.8% (2340/2549)	93.9% (294/313)
Cardiovascular conditions		
New York Heart Association functional class III-IV	94.7% (2413/2548)	94.9% (297/313)
Coronary artery disease	77.8% (1982/2549)	77.0% (241/313)
Prior myocardial infarction	26.0% (660/2536)	29.0% (90/310)
Prior percutaneous coronary intervention	39.8% (1013/2546)	32.4% (101/312)
Prior coronary artery bypass graft	42.7% (1088/2549)	44.7% (140/313)
Cerebrovascular disease	26.3% (656/2499)	27.1% (79/292)
Peripheral vascular disease	42.8% (1078/2520)	43.0% (132/307)
Prior balloon aortic valvuloplasty	23.3% (592/2538)	10.2% (32/313)
Pulmonary hypertension	38.9% (942/2419)	35.4% (95/268)
Major arrhythmia	50.5% (1286/2548)	51.3% (160/312)
Noncardiac conditions		
Coagulopathy	2.3% (58/2542)	3.2% (10/312)
Renal disease	16.6% (423/2548)	19.5% (61/313)
Liver disease	2.7% (68/2546)	2.9% (9/313)
Chronic obstructive pulmonary disease	43.4% (1108/2552)	44.1% (138/313)

Values are presented as median (interquarile range), % (n), or mean ± standard deviation (n). SAVR, Surgical aortic valve replacement; TAVR, transcatheter aortic valve replacement; CA, continued access.

(Figure 2, C and D). These results were consistent among the subgroups of patients undergoing TF and TA access. At 1 year, the calibration was better with STS score compared with LES score for the overall TAVR group (Table 3 and Figure 2, E and F).

Risk Score Performance Characteristics of Patients Undergoing SAVR (n = 313)

Even among this surgical cohort, the mean LES score was 3 times higher than the mean STS score (29.2% ± 15.2% vs 11.7% ± 3.4%), and as such, the observed/predicted 30-day mortality ratio was 0.35 for the LES score

and 0.89 for the STS score. There was a weak ($Y = 0.0597x + 9.7613$; $R = 0.2556$; $r^2 = 0.06$) relationship between the LES and STS scores.

Both were poor discriminators of 30-day/in-hospital outcomes. At 1 year, the discrimination of both STS and LES scores remained poor with a marginally better performance for STS score (Table 2).

Both risk scores demonstrated significant differences between predicted and observed mortalities throughout the entire range of mortality risk. We found better calibration of the STS model but poor calibration of the LES model for the 30-day outcomes. At 1 year the calibration of STS

TABLE 2. Receiver operating characteristics of Society of Thoracic Surgeons (STS) and logistic European System for Cardiac Operative Risk Evaluation (LES) surgical risk scores

	Original validation cohorts	Transcatheter aortic valve replacement (n = 2446)		Surgical aortic valve replacement (n = 307)	
		30-day/in-hospital mortality	1-year mortality	30-day/in-hospital mortality	1-year mortality
STS score	0.78	0.60 (0.55-0.65)	0.56 (0.53-0.58)	0.58 (0.48-0.69)	0.61 (0.54-0.68)
LES score	0.76	0.53 (0.48-0.57)	0.56 (0.53-0.59)	0.58 (0.48-0.67)	0.51 (0.44-0.59)
P value (STS score vs LES)		.024	.83	.93	.042

Values are presented as area under the receiver operating characteristic curve (95% confidence interval). STS, Society of Thoracic Surgeons; LES, logistic European System for Cardiac Operative Risk Evaluation.

TABLE 3. Summary of calibration characteristics* of Society of Thoracic Surgeons (STS) and logistic European System for Cardiac Operative Risk Evaluation (LES) surgical risk scores

	Transcatheter aortic valve replacement (n = 2446)		Surgical aortic valve replacement (n = 307)	
	30-day/in-hospital mortality	1-year mortality	30-day/in-hospital mortality	1-year mortality
STS score	−0.0016 (0.57)	0.084 (1.03)	−0.0322 (1.19)	−0.0368 (2.46)
LES score	0.0552 (0.031)	0.1535 (0.18)	0.1602 (−0.18)	0.2158 (0.11)

Values are presented as intercept (slope). STS, Society of Thoracic Surgeons; LES, logistic European System for Cardiac Operative Risk Evaluation. *Ideal calibration line: $Y = X$, slope = 1, intercept = 0.

score and LES score was poor (Table 3 and Appendix Figure E1).

DISCUSSION

The primary findings of our analysis of risk score performance in the PARTNER I trial and continued access patients were the STS and LES surgical risk scores were poorly correlated and overestimated 30-day/in-hospital mortality of patients undergoing TAVR, both scores were poor discriminators of 30-day/in-hospital mortality as well as 1-year mortality of patients undergoing TAVR, the calibration of these scores within the TAVR and SAVR populations was better for STS score than for LES score, and the calibration of the STS score was better among the patients undergoing SAVR than those undergoing TAVR at 30 days in our analysis. Overall, these data highlight the need for TAVR-specific risk models to optimize patient selection.

The use of the existing surgical risk scores among the cohort of high-risk patients considered for TAVR is controversial. In part due to the absence of adequate risk stratification paradigms, calculation of STS score (or LES score) is routine in the evaluation of patients undergoing evaluation for TAVR. In fact, STS predicted risk of mortality was a key inclusion requirement of the PARTNER trial. Mirroring this inclusion criterion, the Centers for Medicare and Medicaid Services have incorporated the STS-predicted risk of mortality as a benchmark for Medicare TAVR reimbursement. STS score and LES score predicted mortality have thus become an important consideration in the counseling and decision making of patients with AS. As a result, the outcomes of patients undergoing TAVR have been compared with STS score and LES score predicted morbidity and mortality in published literature.¹¹

In the PARTNER experience, both the LES and STS scores overestimated 30-day/in-hospital TAVR mortality with observed to expected mortality ratios of 0.15 and 0.57, respectively. The LES and STS scores also overestimated 30-day SAVR mortality with observed to expected ratios of 0.35 and 0.89, respectively. Additionally, we found a large discrepancy between the STS and LES scores (both prospectively calculated) for both the TAVR and SAVR cohorts. A similar large discrepancy was described by Piazza et al,¹¹ between LES (prospectively calculated) and STS (retrospectively calculated) predicted risks of

mortality when applied to patients undergoing CoreValve implantation. LES is well known to overestimate mortality risk among high-risk patients and octogenarians undergoing SAVR.^{12–15} The derivation of the LES was based on a dataset composed chiefly of coronary surgery patients (approximately 60%), and therefore may be less well adapted than dedicated valve surgery models for prediction of operative mortality in patients undergoing valve surgery.^{16–23} The LES score was not designed to predict mortality in patients undergoing isolated SAVR and has been demonstrated to overestimate mortality in high-risk patients undergoing SAVR by a factor of 3.⁸ The European System for Cardiac Operative Risk Evaluation II score takes into account that cardiac surgical mortality has significantly decreased in the past 15 years despite the inclusion of older and sicker patients as surgical candidates. This score appears to be well calibrated (actual mortality, 4.18%; predicted mortality, 3.95%) and appeared to have good overall discriminatory capacity with an AUC of 0.8095 when applied to a traditional risk surgical population.²⁴ Although the European System for Cardiac Operative Risk Evaluation II model has replaced the logistic or additive models, the latter has been extensively applied to the existing TAVR literature and therefore, while increasingly of historical importance, the performance characteristics of this model remain clinically important to assess. In a recent small study of patients undergoing TAVR, LES and the STS scores were found to be better calibrated than the LES but have moderate discrimination for predicting 30-day mortality after TAVR.²⁵ The application of the European System for Cardiac Operative Risk Evaluation II score to patients undergoing TAVR remains to be validated. The short-term nature of our study still leaves in question the durability of the prosthesis and, therefore, application in patients with more than short- to intermediate-term life expectancy should be limited until longer-term experience reveals acceptable performance.

The STS and LES algorithms not only differ in degree of predicted mortality, but also identify different patients at the highest risk for SAVR.¹³ Dewey et al¹³ previously reported underestimation of mortality by the STS score (an observed to expected mortality ratio of 1.41) and overestimation by the LES score in high-risk patients undergoing SAVR (observed to expected mortality ratio of 0.31) using the

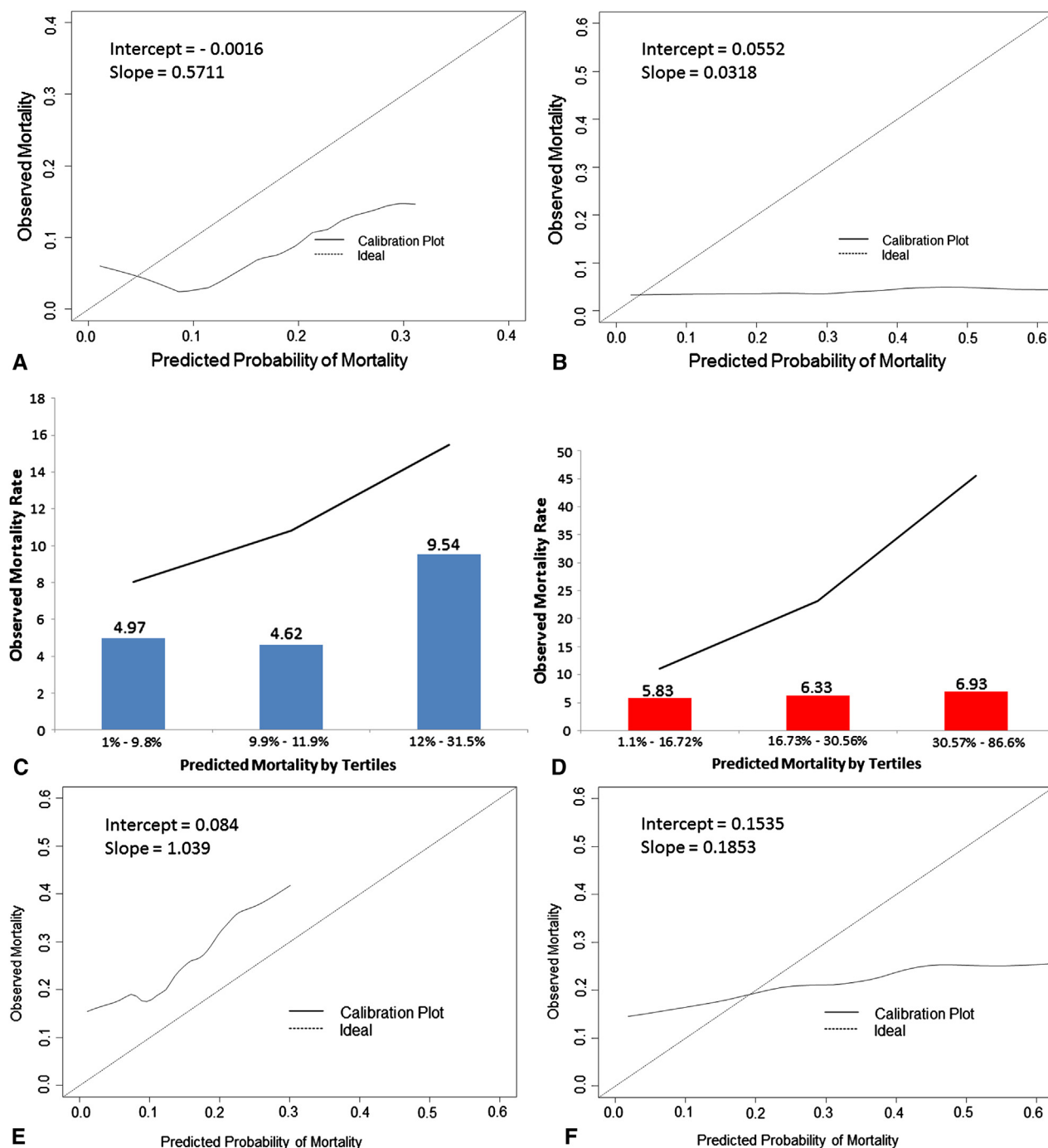


FIGURE 2. Patients undergoing transcatheter aortic valve replacement. A, Calibration plots of the Society of Thoracic Surgeons (STS) scores for 30-day/in-hospital mortality. Perfect calibration is represented by the *dashed line* whereby observed equals predicted probability of mortality ($X = Y$, slope = 1, intercept = 0). B, Calibration plots of the logistic European System for Cardiac Operative Risk Evaluation (LES) score for 30-day/in-hospital mortality. Perfect calibration is represented by the *dashed line* whereby observed equals predicted probability of mortality ($X = Y$, slope = 1, intercept = 0). C, Calibration plots of the STS scores for 30-day/in-hospital mortality analyzed by tertiles. Perfect calibration is represented by the *dashed line* whereby observed equals predicted probability of mortality ($X = Y$, slope = 1, intercept = 0). D, Calibration plots of the LES scores for 30-day/in-hospital mortality analyzed by tertiles. Perfect calibration is represented by the *dashed line* whereby observed equals predicted probability of mortality ($X = Y$, slope = 1, intercept = 0). E, Calibration plots for the STS scores for 1-year mortality. Perfect calibration is represented by the *dashed line* whereby observed equals predicted probability of mortality ($X = Y$, slope = 1, intercept = 0). F, Calibration plots for the LES for 1-year mortality. Perfect calibration is represented by the *dashed line* whereby observed equals predicted probability of mortality ($X = Y$, slope = 1, intercept = 0).

LES score, among 64 high-risk aortic valve replacement patients. A general shortcoming of all surgical risk algorithms, including the STS and LES scores, is the omission of several measurable and immeasurable risk factors known to influence patient selection and mortality.^{7,16-19,26} For example, both models fail to include porcelain aorta, mediastinal radiation, malnutrition, and patient frailty. This should in theory result in an underestimation of mortality, but in our analysis both risk models in fact overestimated mortality.

To guide patient selection, prognostic models must have good discrimination (ie, ability to correctly classify patients as dead or alive) and calibration (ie, comparison between observed and predicted mortality) not only at short-term follow-up but also at longer-term follow-up.²⁷ This is especially true in patients undergoing TAVR where a substantial residual mortality remains following successful TAVR.^{3,4} We found suboptimal discriminatory power of both the STS and LES risk scores for both periprocedural and 1-year mortality among both TF and TA cohorts. For the SAVR population, STS performed poorly and similar to the LES in the short term but performed better than LES at 1 year. The calibration was good for the STS score but poor for the LES score for patients undergoing both TF and TA TAVR as well as for patients undergoing SAVR at both 30-day and 1-year follow-up. In a previous analysis of patients undergoing TAVR using the CoreValve device, the STS score outperformed the LES score, but both LES and STS score had suboptimal discriminatory power and calibration.¹¹ A recently published smaller study of patients undergoing TAVR (n = 250) found that the STS score is better calibrated than the LES score but has moderate discrimination for predicting 30-day mortality after TAVR.²⁵ This lack of discrimination and calibration is not surprising given the extrapolation of surgical risk scores to a high-risk population of TAVR patients underrepresented in the derivation and/or validation datasets from which these scores were derived. This accounts for some of the overestimation of risk seen with many models. Additionally, the risk scores cannot account for the risk of random events and their effect on outcomes, such as operator error during TAVR. Furthermore, given that both surgical risk scores greatly overestimate the actual mortality of patients undergoing TAVR, complacency may arise when these are used as benchmarks of performance. Consistent with our findings, the recently published European Society of Cardiology/European Association for Cardiothoracic Surgery 2012 guidelines point out the lack of accuracy of the STS and LES scores and recommend the development of dedicated risk scores for transcatheter heart valve interventions and in the interim recommend the evaluation and prognostication of patients by a dedicated heart team.²⁸

Limitations

Our analysis presents a secondary analysis of the data collected as part of the PARTNER I randomized controlled trial and continued access registries and therefore should be considered hypothesis-generating. The analyzed patient population was part of the initial North American TAVR experience and the risk profile may not be representative of current high-risk profile patients being screened for TAVR. A larger sample size may better guide the development of an optimal TAVR-specific risk score. Although it would be of interest to determine the performance characteristics of the European System for Cardiac Operative Risk Evaluation II score as applied to the PARTNER dataset, this was not feasible because both STS and LES scores were calculated prospectively at the site level during the screening process when patient eligibility was determined. The calculated scores were then entered into the PARTNER case report form. As such, the individual components contributing to the calculation of each risk score were not systematically captured as part of the case report form. Therefore, several variables used by the European System for Cardiac Operative Risk Evaluation II algorithm (<http://euroscore.org/calc.html>) are not directly available in the PARTNER database. Rather than perform multiple imputations to derive approximated versions of these variables in each patient (which would introduce significant inaccuracy), we chose not to perform these analyses recognizing the inherent limitations in such an approach.

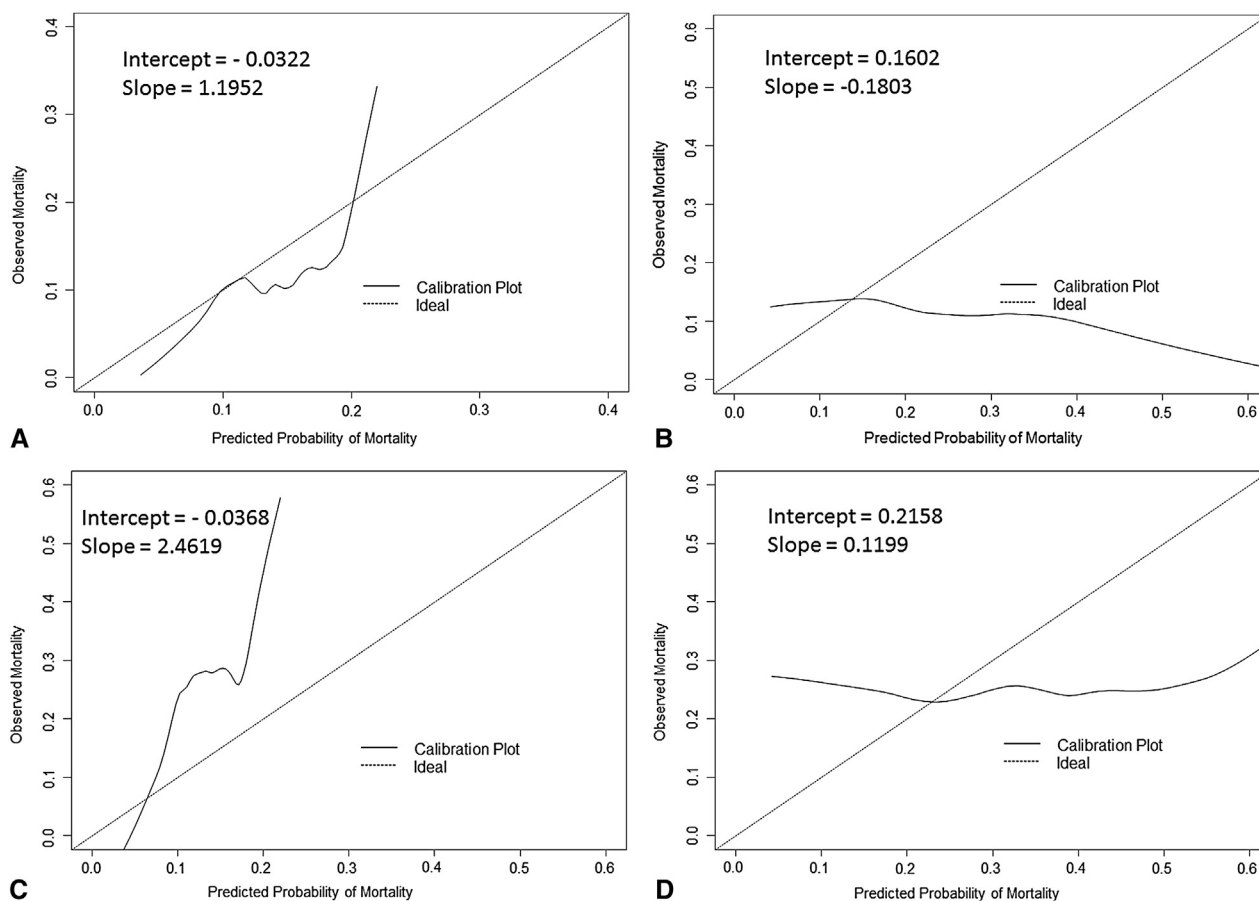
CONCLUSIONS

Existing surgical risk scores are weakly correlated and poor discriminators of 30 day/in-hospital and 1-year outcomes of TAVR and high-risk SAVR, but exhibit some degree of calibration to these outcomes (specifically STS score). Validated TAVR-specific risk models need to be developed to optimize patient selection and improve clinical outcomes. Until such validated TAVR-specific risk scores are available, clinical judgment integrating input from a heart team should guide clinical decision making. Additionally, recalibration of the risk scores, especially LES score, may be warranted to better predict outcomes of high-risk patients undergoing SAVR.

References

1. Smith CR, Leon MB, Mack MJ, Miller DC, Moses JW, Svensson LG, et al. Transcatheter versus surgical aortic-valve replacement in high-risk patients. *N Engl J Med*. 2011;364:2187-98.
2. Leon MB, Smith CR, Mack M, Miller DC, Moses JW, Svensson LG, et al. Transcatheter aortic-valve implantation for aortic stenosis in patients who cannot undergo surgery. *N Engl J Med*. 2010;363:1597-607.
3. Makkar RR, Fontana GP, Jilaihawi H, Kapadia S, Pichard AD, Douglas PS, et al. Transcatheter aortic-valve replacement for inoperable severe aortic stenosis. *N Engl J Med*. 2012;366:1696-704.
4. Kodali SK, Williams MR, Smith CR, Svensson LG, Webb JG, Makkar RR, et al. Two-year outcomes after transcatheter or surgical aortic-valve replacement. *N Engl J Med*. 2012;366:1686-95.

5. Thourani VH. Three-year outcomes after transcatheter or surgical aortic valve replacement in high-risk patients with severe aortic stenosis. Presented at: American College of Cardiology Annual Scientific Sessions 2013, San Francisco, California, March 11, 2013.
6. Shahian DM, Edwards FH. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: introduction. *Ann Thorac Surg.* 2009;88(1 Suppl):S1.
7. Nashef SA, Roques F, Hammill BG, Peterson ED, Michel P, Grover FL, et al. Validation of European System for Cardiac Operative Risk Evaluation (EuroSCORE) in North American cardiac surgery. *Eur J Cardiothorac Surg.* 2002;22:101-5.
8. Mack MJ. Risk scores for predicting outcomes in valvular heart disease: how useful? *Curr Cardiol Rep.* 2011;13:107-12.
9. Zou KH, O'Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation.* 2007;115:654-7.
10. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation.* 2007;115:928-35.
11. Piazza N, Wenaweser P, van Gameren M, Pilgrim T, Tzikas A, Otten A, et al. Relationship between the logistic EuroSCORE and the Society of Thoracic Surgeons Predicted Risk of Mortality score in patients implanted with the CoreValve ReValving system—a Bern-Rotterdam Study. *Am Heart J.* 2010;159:323-9.
12. Brown ML, Schaff HV, Sarano ME, Li Z, Sundt TM, Dearani JA, et al. Is the European System for Cardiac Operative Risk Evaluation model valid for estimating the operative risk of patients considered for percutaneous aortic valve replacement? *J Thorac Cardiovasc Surg.* 2008;136:566-71.
13. Dewey TM, Brown D, Ryan WH, Herbert MA, Prince SL, Mack MJ. Reliability of risk algorithms in predicting early and late operative outcomes in high-risk patients undergoing aortic valve replacement. *J Thorac Cardiovasc Surg.* 2008;135:180-7.
14. Osswald BR, Gegouskov V, Badowski-Zyla D, Tochtermann U, Thomas G, Hagl S, et al. Overestimation of aortic valve replacement risk by EuroSCORE: implications for percutaneous valve replacement. *Eur Heart J.* 2009;30:74-80.
15. Leontyev S, Walther T, Borger MA, Lehmann S, Funkat AK, Rastan A, et al. Aortic valve replacement in octogenarians: utility of risk stratification with EuroSCORE. *Ann Thorac Surg.* 2009;87:1440-5.
16. O'Brien SM, Shahian DM, Filardo G, Ferraris VA, Haan CK, Rich JB, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 2—isolated valve surgery. *Ann Thorac Surg.* 2009;88(1 Suppl):S23-42.
17. Hannan EL, Wu C, Bennett EV, Carlson RE, Culliford AT, Gold JP, et al. Risk index for predicting in-hospital mortality for cardiac valve surgery. *Ann Thorac Surg.* 2007;83:921-9.
18. Kuduvali M, Grayson AD, Au J, Grotte G, Bridgewater B, Fabri BM. A multi-centre additive and logistic risk model for in-hospital mortality following aortic valve replacement. *Eur J Cardiothorac Surg.* 2007;31:607-13.
19. Nowicki ER, Birkmeyer NJ, Weintraub RW, Leavitt BJ, Sanders JH, Dacey LJ, et al. Multivariable prediction of in-hospital mortality associated with aortic and mitral valve surgery in Northern New England. *Ann Thorac Surg.* 2004;77:1966-77.
20. Ambler G, Omar RZ, Royston P, Kinsman R, Keogh BE, Taylor KM. Generic, simple risk stratification model for heart valve surgery. *Circulation.* 2005;112:224-31.
21. Florath I, Rosendahl UP, Mortasawi A, Bauer SF, Dalladaku F, Ennker IC, et al. Current determinants of operative mortality in 1400 patients requiring aortic valve replacement. *Ann Thorac Surg.* 2003;76:75-83.
22. Gardner SC, Grunwald GK, Rumsfeld JS, Cleveland JC Jr, Schooley LM, Gao D, et al. Comparison of short-term mortality risk factors for valve replacement versus coronary artery bypass graft surgery. *Ann Thorac Surg.* 2004;77:549-56.
23. Jin R, Grunkemeier GL, Starr A. Validation and refinement of mortality risk models for heart valve surgery. *Ann Thorac Surg.* 2005;80:471-9.
24. Nashef SA, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR, et al. EuroSCORE II. *Eur J Cardiothorac Surg.* 2012;41:734-44; discussion 44-5.
25. Durand E, Borz B, Godin M, Tron C, Litzler PY, Bessou JP, et al. Performance analysis of EuroSCORE II compared to the original logistic EuroSCORE and STS scores for predicting 30-day mortality after transcatheter aortic valve replacement. *Am J Cardiol.* 2013;111:891-7.
26. Rankin JS, Hammill BG, Ferguson TB Jr, Glower DD, O'Brien SM, DeLong ER, et al. Determinants of operative mortality in valvular heart surgery. *J Thorac Cardiovasc Surg.* 2006;131:547-57.
27. van Gameren M, Piazza N, Bogers AJ, Takkenberg JJ, Kappetein AP. How to assess risks of valve surgery: quality, implementation and future of risk models. *Heart.* 2009;95:1958-63.
28. Joint Task Force on the Management of Valvular Heart Disease of the European Society of C, European Association for Cardio-Thoracic S, Vahanian A, Alfieri O, Andreotti F, Antunes MJ, et al. Guidelines on the management of valvular heart disease (version 2012). *Euro Heart J.* 2012;33:2451-96.



APPENDIX FIGURE E1. Patients undergoing surgical aortic valve replacement. A, Calibration plots for the Society of Thoracic Surgeons (STS) scores for 30-day/in-hospital mortality. Perfect calibration is represented by the *dashed line* whereby observed equals predicted probability of mortality ($X = Y$, slope = 1, intercept = 0). B, Calibration plots for the logistic European System for Cardiac Operative Risk Evaluation (LES) score for 30-day/in-hospital mortality. Perfect calibration is represented by the *dashed line* whereby observed equals predicted probability of mortality ($X = Y$, slope = 1, intercept = 0). C, Calibration plots for the STS scores for 1-year mortality. Perfect calibration is represented by the *dashed line* whereby observed equals predicted probability of mortality ($X = Y$, slope = 1, intercept = 0). D, Calibration plots for LES for 1-year mortality. Perfect calibration is represented by the *dashed line* whereby observed equals predicted probability of mortality ($X = Y$, slope = 1, intercept = 0).